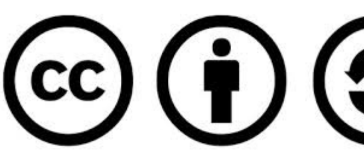


Módulo 3: Gestión de colecciones digitales



Material bajo una Licencia Atribución-Compartir Igual de Creative Commons
https://creativecommons.org/licenses/by-sa/4.0/

Un proyecto de digitalización no termina con la transformación de obras analógicas a digitales. Ese es apenas el comienzo. Tener un montón de información guardada en las carpetas de un disco duro no es el objetivo primordial del proyecto. Este acopio de material digitalizado no constituye una colección organizada de objetos digitales, para lo cual tenemos que dar unos pasos más, que se verán en este módulo.

Estos pasos implican seleccionar cuáles objetos digitalizados vamos a poner a disposición, clasificarlos en colecciones y subcolecciones según criterios que hay que definir, catalogarlos y añadirles descripciones y metadatos que les den contexto, para agregar por fin todo esto en un repositorio online que habilite el acceso al público.

¿Qué es un objeto digital?

Como resultado de la digitalización obtendremos una serie de documentos electrónicos, esto es, una variedad de objetos de información expresados en código digital con distintas propiedades que tienen que ver, entre otras cosas, con el tipo de archivo, formato, contenido, tamaño en bytes, apariencia, estructura, etc.

De cada obra digitalizada se van a generar varios objetos digitales. Por ejemplo, de un libro digitalizado podemos archivar una imagen digital por cada página, y cada una de ellas será un objeto digital independiente. A su vez, si reunimos todas esas imágenes y las transformamos en un documento PDF unificado, obtendremos otro objeto digital, con otras características y propiedades. Otro ejemplo: al digitalizar una película, podemos generar archivos de diversos tipos según su codificación (MPEG, FLV, etc.), y también podríamos generar fotogramas sueltos en una variedad de formatos de imagen fija. Todos serán objetos digitales distintos e independientes; algunos son más adecuados como copias de preservación y otros más aptos para poner a disposición del público en Internet.

Los objetos digitales pueden ser nacidos de soportes físicos (los que se obtienen como resultado de procesos de digitalización) o "nacidos digitales" (los que se crean con distintas herramientas informáticas sin ser la expresión de un objeto "original" del mundo físico). Sin embargo, el uso de la palabra "objeto" es apenas un eufemismo, porque los objetos digitales son representaciones momentáneas que se nos aparecen como objetos al momento de acceder a ellos. Lo que sucede cuando accedemos a un objeto digital no es más que una operación tecnológica que emula la aparición de un objeto que en definitiva es virtual.

Aunque parezca contraintuitivo, los objetos digitales pueden ser igual o más vulnerables que los objetos físicos ante posibles pérdidas o alteraciones. Por eso existe toda un área de trabajo relacionada con la preservación digital, que muchas veces implica la migración de un objeto digital a distintos soportes de información, o su transformación de un formato a otro, para combatir los efectos de la obsolescencia tecnológica.

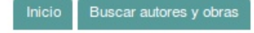
¿Qué es una colección digital?

Para que sean accesibles y fáciles de encontrar e identificar, los objetos digitales suelen organizarse en colecciones digitales. Una colección o biblioteca digital es una selección estructurada de objetos

digitales. El resultado final del proyecto de digitalización será la puesta en línea de una o de varias colecciones en un repositorio accesible al público en Internet. En cada colección, los materiales deben estar bien catalogados y organizados para ser fácilmente recuperables mediante buscadores internos y externos, así como a través de una navegación amigable y bien estructurada.

En el siguiente ejemplo, vemos una de las colecciones digitales del portal autores.uy en la que se reúnen las obras narrativas de Felisberto Hernández. Vemos que la colección tiene una URL fija (http://autores.uy/coleccion/635) que permite hacer un enlace permamente a ella desde cualquier otra página web. Además del nombre y descripción general de la colección, aparecen las obras listadas con su respectivo título, año de publicación, tipo de obra y una imagen asociada, en este caso, una miniatura de la portada. Haciendo clic en el título de la obra, accedemos a ella y podemos descargarla.











Creative Commons Uruguay / Fundación Felisberto Hernandez

Obras publicadas por Creative Commons Uruguay con los textos revisados por la Fundación Felisberto Hernández en el marco de un convenio de colaboración entre ambas entidades.

Título	Autores	Año de publicación	Tipo	Imagen
El caballo perdido	Felisberto Hernández	2016	Libro	
Fulano de tal	Felisberto Hernández	2016	Libro	
La cara de Ana	Felisberto Hernández	2016	Libro	

Dentro de una colección digital, cada obra debe tener también una URL o dirección web única y permanente, y si hay varios archivos asociados a la obra (por ejemplo, libros en distintos formatos o fotos en diferentes tamaños), cada uno de esos archivos también tendrá una URL fija.

Como vemos en este ejemplo (http://autores.uy/obra/8828), la obra generalmente se presenta en una ficha que contextualiza y muestra datos sobre ella. Algunos de estos datos vinculan a la obra con otras obras y se presentan en la forma de hipervínculos. Por ejemplo: el autor no es sólo un dato, sino también un vínculo entre las distintas obras del mismo autor.

Los vínculos internos en una colección son fundamentales, así como los enlaces externos que permiten llegar a otras colecciones en donde se encuentre la misma obra y otras. En este ejemplo, la obra de Felisberto Hernández “El caballo perdido” también se encuentra en Internet Archive y en Wikisource, dos repositorios externos que mejoran la visibilidad del objeto y permiten al público acceder a otras versiones del mismo (por ejemplo, hipertexto, audiolibro, etc.).



En el ítem tal como se lo encuentra en Internet Archive, podemos observar que hay varios objetos digitales que el usuario puede seleccionar (EPUB, KINDLE, DAISY, entre otras opciones de descarga).

¿Qué es un repositorio digital?

Una colección digital puede ser simplemente un listado de obras agrupadas en una página web siguiendo algún orden simple, por ejemplo, alfabético. Pero actualmente esto no satisface todas las necesidades de los usuarios a la hora de encontrar obras de su interés, y evidentemente puede ser muy difícil de gestionar si la colección cuenta con muchas obras.

Las colecciones digitales se gestionan mediante un sistema informático denominado repositorio digital que permite almacenar, ordenar, catalogar, poner a disposición del público y encontrar los objetos digitales. Mediante este sistema, los archivos digitales se cargan mediante una aplicación web una única vez, son adjudicados a colecciones, se le agrega la información y los metadatos correspondientes y se publican.

Así, cada obra y cada uno de los objetos que la representan en el repositorio, tendrán una única URL. Asimismo, la obra quedará asociada a distintas categorías de información, como autor, año, género, colección, etc. Como veremos más adelante, los repositorios por lo general incorporan un criterio estándar para asignar metadatos, por lo que no hace falta crear desde cero un sistema de catalogación propio.

Software libre para gestionar las colecciones

En la actualidad existen muchas opciones de software para crear repositorios y gestionar colecciones. Nuestra recomendación es elegir software libre que no nos haga dependientes de licencias onerosas. Para ello, es necesario contar con un servidor web en donde podamos instalar el repositorio y gestionar todos los recursos, tal como se hace con el sitio web de la institución. Si es posible, lo ideal es que los servidores web que alojan el repositorio estén en territorio nacional o en la ubicación geográfica más cercana a la mayoría de los usuarios, para que el acceso sea lo más ágil posible. Contar con los servidores del repositorio en territorio nacional brinda también seguridad jurídica, dado que asegura que la información guardada allí quedará sujeta a la legislación del país en el que reside la institución.

Para construir el repositorio se puede optar a grandes rasgos entre:

- **Utilizar un sistema de gestión de contenidos web genérico como, como WordPress o Drupal**, que proveen las herramientas básicas de almacenamiento de archivos, gestión de contenidos, búsquedas, etc. La ventaja es que son programas ampliamente usados, con

millones de usuarios en todo el mundo, comunidades que brindan soporte y actualizaciones permanentes. La desventaja es que para algunas funciones más avanzadas hay que agregar extensiones que no vienen por defecto, para lo cual hay que contar con algunas habilidades de programación. Por ejemplo, para convertir un sitio WordPress en un repositorio, se podría agregar la extensión **Tanaican**, que resuelve una gran parte de las necesidades de gestión de colecciones digitales.

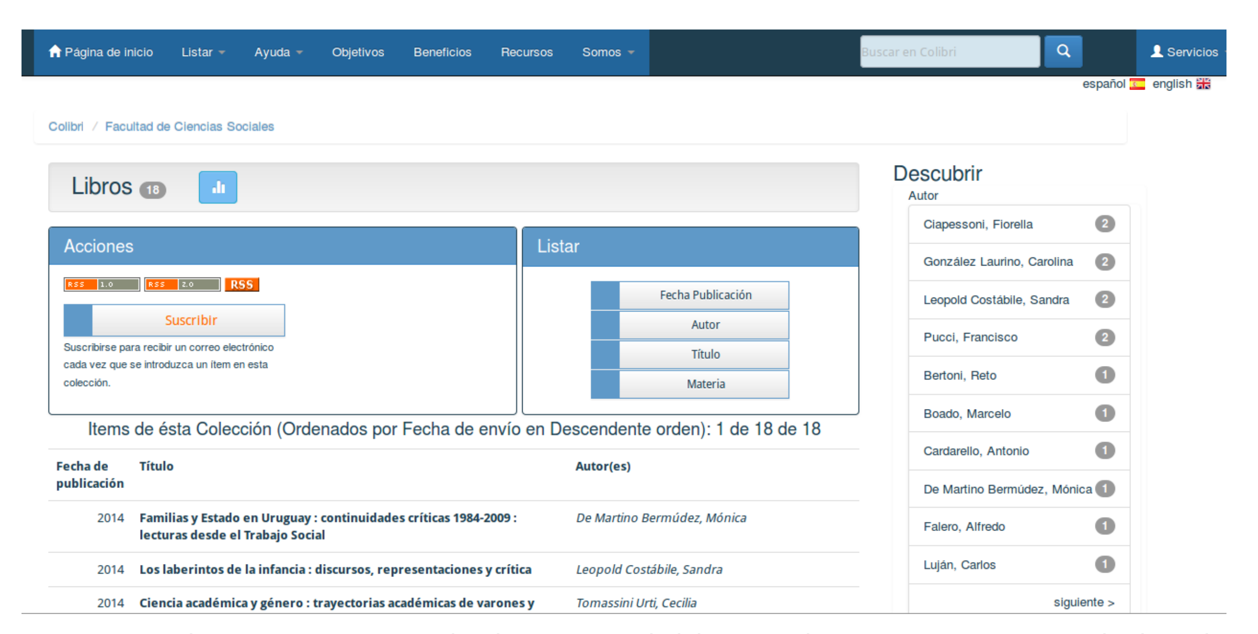
- **Utilizar un sistema de gestión específico para repositorios digitales, como Omeka o DSpace.** Estas herramientas ya vienen preparadas con todas las funciones para cargar, catalogar y visualizar colecciones digitales.



Exhibición digital “Art Nouveau” de Europeana, gestionada con Omeka.

En este ejemplo, vemos una exhibición online gestionada con Omeka. Las exhibiciones son una función de esta herramienta que permite que, una vez alojados y catalogados los objetos en el repositorio, podamos utilizarlos tantas veces como queramos de distintas maneras. Así, la fotografía de un edificio puede integrar esta exhibición de Art Nouveau y más adelante formar parte de otra muestra virtual sobre arquitectura de principios del siglo XX, sin necesidad de subir el objeto digital dos veces.

El siguiente ejemplo corresponde a un repositorio académico gestionado con DSpace. Vemos que una forma de explorar las obras es navegar por los distintos autores de las mismas. Con esta herramienta, cuando se carga una nueva obra, cargamos también los datos correspondientes al autor. Pero será necesario hacerlo solamente la primera vez, ya que las próximas obras del mismo autor no requerirán volver a tipear el nombre del mismo; el autor ya está creado en la base de datos y solamente hará falta asignar nuevas obras a este autor ya existente.



DSpace aplicado en el repositorio Colibrí de la Universidad de la República, Uruguay. Colección de libros de la Facultad de Ciencias Sociales: https://www.colibri.udelar.edu.uy/handle/123456789/4560

Para explorar y comprender este tipo de herramientas, recomendamos comenzar creando una cuenta gratuita en el portal de Omeka: http://www.omeka.net/. Esto nos permitirá evaluar la herramienta y más tarde tendremos la opción de instalarla en un servidor propio.

En el siguiente tutorial encontrarás una guía muy completa para utilizar este software: http://www.rubenalcaraz.es/manual-omeka/objetos-digitales-omeka.html

Repositorios adicionales

Más allá de tomar la decisión de crear un repositorio digital propio, es recomendable cargar las obras en los principales repositorios web de referencia. Son los sitios en los que los usuarios entran a hacer búsquedas y explorar contenidos, por lo que servirán como una "vidriera" y puerta de entrada a nuestras colecciones.

Estamos hablando de un rango amplio de sitios, desde los servicios online de empresas privadas, como Flickr (fotografías), Soundcloud (audios), YouTube y Vimeo (videos) hasta proyectos de patrimonio digital sin fines de lucro, como Internet Archive, Wikimedia Commons o Wikisource.

Como mínimo, recomendamos abrir una cuenta en Internet Archive, que permite subir todo tipo de archivos, sin límites de tamaño y cantidad, pudiendo crear una ficha muy completa para cada obra. Es una herramienta diseñada para archivos y bibliotecas que toma en cuenta las necesidades de almacenamiento y puesta a disposición del patrimonio. Es una iniciativa sin fines de lucro, no presenta publicidad en su sitio y no le exige al usuario loguearse para poder acceder y descargar las obras. Es necesario recordar, eso sí, que Internet Archive solamente admite obras en dominio público, bajo licencias Creative Commons o de las cuales se cuente con autorización explícita del titular.

La importancia de los estándares de almacenamiento y catalogación

Para un sistema informático el significado de un objeto digital es totalmente indiferente. La máquina no puede entender por sí misma la diferencia entre un manuscrito medieval y una película documental de la segunda guerra mundial. Si le decimos a la máquina que el libro es de otra época o la película es de otro tema, la podremos “engañar” fácilmente, porque no distingue significados. Entonces, ¿cómo hace más tarde el usuario para consultar en un sistema informático materiales sobre la segunda guerra, si la máquina no sabe cuáles son?

La forma de hacer que la máquina “reconozca” las obras que posiblemente está buscando el usuario, es catalogando correctamente los objetos digitales, siguiendo siempre un mismo estándar. Esto implica que cuando subimos un archivo digital al repositorio, vamos a cargar también una serie de datos sobre el mismo: estos son los denominados “metadatos”. Como mínimo, un título, pero esto no es suficiente. Podemos agregar autor, tema, fecha, estado de derechos de autor, formato y otros aspectos que describen al objeto. Esto permitirá a la máquina encontrar y recuperar los objetos, y facilitará que el público, al hallarlos, conozca más sobre lo que ha encontrado y sobre su contexto. Con algunos de estos datos los repositorios generan links automáticos que hacen que los objetos se enlacen unos con otros, para que el usuario pueda navegar por la colección.

Hemos dicho que la catalogación debe seguir “un mismo estándar”. Esto significa plegarse a ciertos acuerdos en cuanto al vocabulario utilizado para describir los objetos digitales. Si bien es posible crear una forma *sui generis* de designar y describir nuestros archivos, es un trabajo laborioso y en

definitiva poco útil, ya que los sistemas informáticos tienen incorporados ciertos estándares. Si los ignoramos, hacemos que nuestro archivo resulte incompatible con otros.

Por ejemplo: si ponemos a disposición en nuestro repositorio un archivo musical en MP3 sin descripción, o con una descripción que no sigue los estándares, cuando el usuario lo descarga y lo reproduce en su computadora, su software de reproducción multimedia no va a mostrar correctamente los datos del archivo. El reproductor de MP3 va a leer automáticamente los metadatos e intentará mostrar al usuario el nombre de la canción y del artista en los campos correspondientes. Pero si el archivo está mal catalogado y sus propiedades no están bien descriptas, quizás donde debería figurar el dato del autor, aparecerá el nombre de la institución que digitalizó el audio, o peor aún, el nombre de usuario correspondiente a la computadora en la cual se editó el archivo (este es un error muy común).

Para prevenir este tipo de errores y mantener una colección coherente recomendamos:

- **Al editar el objeto digital, y antes de subirlo al repositorio, editar sus propiedades**. Los principales programas para trabajar con texto, imagen, audio y video, tienen la opción de crear y editar las propiedades de los archivos. Por lo general, se trata de una ficha con algunos campos básicos que podemos completar, como título, autor, palabras clave, etc. Otras propiedades se generan automáticamente, como la fecha de la última modificación, tipo de archivo, tamaño, etc. Todos estos datos "viajan" con el archivo y pueden ser leídos por personas y por máquinas.
- **Incorporar un modelo de metadatos en el repositorio, con una serie de campos estandarizados que sean ampliamente reconocidos**. Un modelo muy utilizado es Dublin Core, que viene incluido en sistemas de gestión como los que vimos (Omeka, DSpace y otros). Al cargar el archivo en el repositorio, veremos una ficha cuyos campos siguen el estándar Dublin Core. Algunos campos se completarán automáticamente con los metadatos que ya tiene el archivo (el título, por ejemplo) y otros los podemos completar en el momento, o corregir los errores que encontremos.

En síntesis, un modelo estándar de metadatos, como Dublin Core, no es ni más ni menos que una convención. Sirve para que cualquier persona y cualquier máquina que tenga incorporado el estándar, sepa identificar datos referentes al autor, palabras clave, tipo de recurso, etc. En total, hay 15 campos o descripciones semánticas en este estándar, cuyas definiciones son acordadas por consenso entre los miembros de la Dublin Core Metadata Initiative. El objetivo de la iniciativa Dublin Core es la interoperabilidad: cuando un mismo modelo de metadatos se mantiene a través

de distintos repositorios, las búsquedas serán más fáciles para los usuarios, la migración de un sistema de repositorio a otro será menos compleja para la institución y la incorporación de una obra existente en un repositorio externo será un paso muy simple. Esencialmente, esto sirve para que el trabajo de generación de información a partir de un objeto se lleve a cabo una sola vez. Por ejemplo: Internet Archive utiliza Dublin Core para estructurar los metadatos. El portal autores.uy, que tiene por objetivo generar un acervo digital de obras uruguayas en dominio público, utiliza el mismo estándar. Cada vez que se encuentra en Internet Archive una obra que puede ser incorporada a la colección, catalogarla resulta mucho más fácil, ya que campos como "Autor o creador" para Internet Archive y para autores.uy significan lo mismo y tienen la misma estructura. La ficha prácticamente se genera sola, por lo que hay que tipear menos, ahorrando tiempo y esfuerzo.